Published online January 22, 2007

Genetic Mapping of Biomass Production in Tetraploid Alfalfa

Joseph G. Robins, Diane Luth, T. Austin Campbell, Gary R. Bauchan, Chunlin He, Donald R. Viands, Julie L. Hansen, and E. Charles Brummer*

ABSTRACT

Biomass production represents a fundamental biological process of both ecological and agricultural significance. The genetic basis of biomass production is unknown but asssumed to be complex. We developed a full sib, F_1 mapping population of autotetraploid *Medicago sativa* (alfalfa) derived from an intersubspecific cross that was known to produce heterosis for biomass production. We evaluated the population for biomass production over several years at three locations (Ames, IA, Nashua, IA, and Ithaca, NY) and concurrently developed a genetic linkage map using restriction fragment length polymorphism (RFLP) and simple sequence repeat (SSR) molecular markers. Transgressive segregants, many of which exhibited high levels of heterosis, were identified in each environment. Despite the complexities of mapping within autotetraploid populations, single-marker analysis of variance identified 41 marker alleles, many on linkage groups 5 and 7, associated with biomass production in at least one of the sampling periods. Seven alleles were associated with biomass production in more than one of the sampling periods. Favorable alleles were contributed by both parents, one of which is from the *M. sativa* subsp. *falcata* germplasm. Thus, increased biomass production alleles can be gleaned from unadapted germplasm. Further, the positive quantitative trait locus (QTL) alleles from the parents are partially complementary, suggesting these loci may play a role in biomass production heterosis.

Despite the agronomic and ecological importance of biomass production, its genetic control has not been well studied. Biomass production is expected to result from the complex interaction of many genes within a variable environmental context, and hence, it is not amenable to simple genetic dissection. Nevertheless, similar complexity has been investigated in numerous QTL mapping experiments to identify genomic regions associated with production of seed or fruit (e.g., Austin and Lee, 1998; Bernacchi et al., 1998). Heterosis, or the superiority of hybrid progeny relative to their parents, has been shown to increase biomass and/or seed yield of various crops, yet the genetic control of heterosis is also unknown in any crop.

Medicago sativa produces highly nutritious biomass under a wide range of environmental conditions. Following removal of aboveground biomass, *M. sativa* regrows from crown and auxiliary buds, resulting in several flushes of biomass production each year. In addition, *M. sativa* plants growing in the temperate parts of the world have the ability to enter physiological dormancy in autumn, enabling them to survive winter and recommence biomass production the following spring. Thus, over a several-year period, biomass production in *M. sativa* will vary depending on both the environmental conditions and the developmental stage of a given plant. Unlike woody species, in which each new seasonal production of biomass extends the previous year's growth, herbaceous perennials like *M. sativa* have all aboveground biomass removed repeatedly so that each growth period replaces, rather than augments, the biomass produced previously. As trees age, some of the genomic regions associated with biomass production appear to change (Wu, 1998; Lerceteau et al., 2001), but the situation for alfalfa is unknown.

Extensive variation among alfalfa populations for biomass production is widely recognized by plant breeders, yet over the past 25 yr, applied alfalfa breeding efforts in the upper Midwest have not increased the biomass yield potential of commercially cultivated alfalfa substantially (Riday and Brummer, 2002). The improved alfalfa yield of newer cultivars relative to older ones appears to be attributable to increased pest resistance rather than to improved yield potential per se (Lamb et al., 2006). Two obvious methods to improve biomass production, neither of which has been extensively exploited in alfalfa, are (i) direct selection specifically for biomass production and (ii) the exploitation of heterosis through hybrid or semihybrid alfalfa cultivars (Brummer, 1999). Biomass yield heterosis has been identified in alfalfa (reviewed in Brummer, 1999; Riday and Brummer, 2002), but hybrid alfalfa cultivars have not been a focus of most breeding programs, with only a few hybrid cultivars commercially available (e.g., Hybriforce 400 from Dairyland Seeds, Inc., West Bend, WI).

The tetrasomic tetraploid genome of *M. sativa*, associated with an allogamous breeding system and a general intolerance of inbreeding, complicates genetic analyses. Although the theory of genetic linkage in autopolyploids is well developed (Haldane, 1930; DeWinton and Haldane, 1931; Mather, 1936), the difficulty of resolving allele dosage and linkage phases limits the information content that can be gathered from most current molecular markers. To avoid these complications, diploid relatives of cultivated polyploids have been mapped in several species, including alfalfa (e.g., Brummer et al., 1993; Kiss et al., 1993; Echt et al., 1994). While diploid mapping avoids the complexities of polysomic inheritance and works well if the synteny across ploidy levels is high, it

Joseph G. Robins, USDA-ARS Forage and Range Research Lab., Logan, UT 84322; Diane Luth, and E. Charles Brummer, Raymond F. Baker Center for Plant Breeding, Iowa State Univ., Ames, IA 50011; E. Charles Brummer current address: Center for Applied Technologies, Univ. of Georgia, Athens, GA 30602; T. Austin Campbell, Gary R. Bauchan, and Chunlin He, USDA-ARS Soybean Genomics and Improvement Lab., Beltsville, MD 20705; Donald R. Viands and Julie L. Hansen, Dept. of Plant Breeding and Genetics, Cornell Univ., Ithaca, NY 14853. Received 3 Nov. 2005. *Corresponding author (brummer@uga.edu).

Published in Crop Sci. 47:1–10 (2007).
Crop Breeding & Genetics
doi:10.2135/cropsci2005.11.0401

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: FWER, familywise error rate; LG, linkage group; QTL, quantitative trait locus; RFLP, restriction fragment length polymorphism; SSR, simple sequence repeat.

might not be useful if the genetic control of a phenotype differs across ploidies. Evidence for differential genetic control across ploidies has been shown by gene expression profiling in yeast (Galitski et al., 1999) and by quantitative genetics in *M. sativa* (Groose et al., 1988). Several mapping studies have been conducted in tetraploid *M. sativa* (Brouwer and Osborn, 1999; Julier et al., 2003; Sledge et al., 2005), but none of these studies has investigated the underlying genetics of biomass production.

The objectives of this experiment were (1) to develop a genetic linkage map of tetraploid *M. sativa*, (2) to identify genomic regions in tetraploid *M. sativa* associated with aboveground biomass production across developmental stages of plants growing in multiple environments, and (3) to infer the loci underlying biomass production heterosis.

MATERIALS AND METHODS

Plant Material

A cross between the autotetraploid genotypes WISFAL-6 × ABI408 was made using vacuum emasculation of the female parent, resulting in a segregating, full sib F_1 population of 200 genotypes. WISFAL-6 is a semi-improved *M. sativa* subsp. *falcata* genotype from the WISFAL germplasm (Bingham, 1993), and ABI408 represents an elite *M. sativa* subsp. *sativa* genotype from ABI Alfalfa, Inc. (Lenexa, KS). These parents were selected for use in mapping because of evidence that their hybrid progeny exhibited significant amounts of heterosis for biomass yield, which was documented in a separate experiment (Riday and Brummer, 2002). This population was also described by Brummer et al. (2000). The 200 F_1 genotypes, the parents, and eight check genotypes were clonally propagated by stem cuttings in the greenhouse at Ames, IA, for use in DNA extraction and for phenotypic analysis.

Genotyping and Genetic Map Construction

Genomic DNA was extracted from leaves using the method of Doyle and Doyle (1990). Extracted DNA was then used for RFLP and SSR analysis. RFLP analysis was performed using the procedures of Brummer et al. (1993) using the restriction enzymes *Eco*RI and *Hin*dIII. Separate sets of population blots were made for each enzyme. Probes came from the following sources: Drs. G. D. Kochert and J. H. Bouton (Botany and Crop and Soil Sciences Depts., Univ. of Georgia, Athens, GA); Dr. T. C. Osborn (Plant Breeding Dept., Univ. of Wisconsin, Madison, WI); Dr. J. J. Volenec (Agronomy Dept., Purdue Univ., W. Lafayette, IN); Dr. S. Arcioni (Istituto di Ricerche sul Miglioramento Genetico delle Piante Foraggere, Perugia, Italy); Dr. B. D. McKersie (Plant Agriculture Dept., Univ. of Guelph, Guelph, ON); and Dr. S. Laberge (Agriculture and Agri-Food Canada, Sainte Foy, QC). Primer sequences for SSR markers were acquired from several sources (Diwan et al., 2000; Thoquet et al., 2002; Julier et al., 2003; Eujayl et al., 2003) (Supplementary Table 1). Simple sequence repeat loci were identified based on the nomenclature used by Diwan et al. (2000), Thoquet et al. (2002), or the corresponding Genbank accession name. The polymerase chain reaction amplifications of SSR markers were based on the method of Diwan et al. (1997) and were detected using either a Licor 4200 DNA Analyzer or an ABI 3100 DNA Analyzer. Data were scored using the AFLP-Quantar (Keygene), Gene Scan, or Genotyper software (ABI). Each allele was scored for its presence or absence in each progeny genotype. Each allele was designated with a letter denoting the parent carrying it, with WISFAL-6 designated as "a," ABI408 as "b," and alleles present in both parents as "c." Multiple alleles produced by a single probe or primer were differentiated based on band size using numbers following the letter designations. A subset of loci that were scored on the ABI 3100 and exhibited distorted segregation was reanalyzed using the Licor 4200; the results were consistent between the two platforms.

The TetraploidMap program suite (Hackett and Luo, 2003) was used to infer the parental dosage (number of copies) of each allele and the most likely parental genotype, to identify possible double reductants, to cluster markers into linkage groups (LGs), and to calculate recombination frequencies and accompanying LOD scores (Supplementary Table 2). A consensus map representing both parental genomes, like that of Julier et al. (2003), was constructed by combining all marker data from both parents in the TetraploidMap analysis. Marker DNA fragments were considered to be allelic if the most likely genotype at a given locus contained the alleles being evaluated; otherwise, the fragments were considered to belong to duplicated loci. Using LOD values ≥ 3.0 and recombination frequencies ≤ 0.30, JoinMap 3.0 (Van Ooijen and Voorrips, 2001) was used to order the codominant marker data within LGs, and MapChart (Voorrips, 2002) was used to draw the resulting LGs. Linkage group numbering corresponds to that of *M. truncatula* (Thoquet et al., 2002; Choi et al., 2004).

The result of the overall analysis was eight consensus LGs representing the basic set of eight *M. sativa* chromosomes. Thus, each of these consensus LGs was based on marker information from the four homologous chromosomes from each parent. To visualize the linkage arrangements of markers on the individual homologs, each consensus LG was individually decomposed into the eight constituent cosegregation groups, or a total of 32 cosegregation groups for each parent (Supplementary Fig. 1). The cosegregation groups were determined by manually analyzing the recombination frequency output from the TetraploidMap program and identifying alleles linked in coupling phase. The cosegregation groups of each consensus LG were determined by systematically analyzing the linkage relationships of each marker with the other markers in the group and placing alleles linked in coupling phase in the same cosegregation group. The vast majority of alleles were placed into cosegregation groups based on coupling-phase linkages with an LOD ≥ 3.0. In a few instances, placement occurred with an LOD < 3.0 or was inferred from repulsion-phase linkages. This most often occurred with alleles that were present in multiple copies in one or both parents. These alleles typically had strong linkage support for one copy to be included in a cosegregation group but with less support for inclusion of the second (or third) copy of the allele into other cosegregation groups. The order of the markers in each cosegregation group was based on their placement on the consensus LG. Because the consensus LG included all the recombination information in the population, it offers a more realistic locus order than would be produced by remapping the individual cosegregation groups, which are more sparsely populated with markers.

Phenotyping

Experimental Design

Field experiments were planted at the Agronomy and Agricultural Engineering Research Farm west of Ames, IA, on 19 May 1998; at the Northeast Research Farm south of Nashua, IA, on 22 May 1998; and at the Snyder 5 east field adjacent to the Game Farm Road Weather Station in Ithaca, NY, on 7–9 June 1999. The plot design at Ames and Nashua was a quadruple α-lattice consisting of 840 total plots (each replication consisted of 15 incomplete blocks each with 14 plots)

and at Ithaca was a randomized complete block design consisting of 4 blocks and 824 total plots. The difference in total plot numbers between IA and NY was due to the loss of six genotypes before transplanting and the inclusion of two additional check cultivars in NY. Plots in IA consisted of five clones of each genotype; in NY, they consisted of seven clones, but only the inner five clones were harvested. In IA, spacings were 30 cm between plants within a plot, 60 cm between plots in the same row, and 75 cm between rows. In NY, spacings were 25 cm between plants within a plot, 60 cm between plots within the same row, and 90 cm between rows. In addition, in NY the experiment was overseeded with red fescue on 9 Sept. 1999 to limit weed competition.

Phenotypic Data Collection

After the initial establishment year, biomass production was measured for 3 yr (1999–2001) in IA and for 2 yr (2000–2001) in NY. Data were collected on three harvests per year (June, July, and September) at each location with the exception of NY in 2000, when excessive rainfall resulted in no data collection in June.

Harvesting consisted of the removal of all aboveground biomass to ~7.5 cm above the soil surface. Harvesting was conducted by hand, using rice sickles, during 1999 in IA and for both years in NY. A flail-type, self-propelled forage harvester (Carter Manufacturing Co., Inc., Brookston, IN) was used in IA during the 2000 and 2001 harvest years. The mass of wet forage from each plot was determined in the field. Random subsamples were taken from the harvested material in each replication, weighed, dried for 4 d at 60°C, and reweighed to compute a dry matter percentage. An average dry matter percentage across the subsamples was used to adjust plot wet mass to a dry matter basis. The number of plants present in each plot was counted after each harvest and plot dry matter production recorded as g plant^{-1}.

Phenotypic Data Analysis

Biomass from each of the three individual harvests per year was summed to produce the yearly biomass production for each plot. To keep this article to manageable length, we only consider total yearly biomass production. The complete model was analyzed with genotype, location, and their interaction being fixed effects, and replications and incomplete blocks being random effects. Data from IA were analyzed to determine the effect of both locations and genotype × environment interactions. Least-square means on a per entry basis for yearly totals were calculated for the combined data from the two IA locations and separately for NY using the MIXED procedure (SAS statistical software, Cary, NC; Littell et al., 1996). Data from IA and NY were not combined because the experiment in NY began 1 yr after the IA experiment, and the first harvest from NY (in 2000) was lost due to excessive rainfall. Due to the perennial nature of alfalfa and to heterogeneous error estimates between years, growth in each year was treated as a separate and distinct trait, which allowed more thorough investigation of changes in the genetic basis of biomass production from year to year as the plants aged.

Variance components and heritabilities with their standard errors were computed from an all random effects model (Holland et al., 2003) with the parents and check genotypes removed. The consideration of genotype as a random effect, as opposed to the previous fixed effect assumption, was necessary to calculate the variance associated with this effect. The genetic variation includes all higher-order intralocus and epistatic interactions present in a tetrasomic tetraploid (Rumbaugh et al., 1988). Because the phenotypic data came from a single full sib population with no ability to partition the additive genetic variance from the dominance genetic variance, broad-sense heritability estimates were generated (Holland et al., 2003). Phenotypic and genetic correlations were calculated using the MIXED and IML procedures (Holland et al., 2006), which also produced the corresponding standard errors. Correlations between years were calculated for IA and NY separately. All discussions of statistical significance were based on the 5% probability level, unless otherwise noted.

Marker–Phenotype Associations (QTL Analysis)

Marker–phenotype associations were calculated using single-marker analysis of variance with the GLM procedure of SAS. The least-square means of individuals containing an allele were contrasted to those of individuals not containing the allele. Because this is the first study examining potential QTLs for biomass production in alfalfa, we were more concerned about identifying putative genomic regions associated with the trait than about false positive associations. For this reason, we set the cutoff value for declaring an association between a marker allele and a phenotype at $\alpha = 0.01$. Although this level will lead to some spurious associations, those markers identified in more than one environment would be unlikely to arise by chance. To identify associations based on a familywise error rate (FWER), we used nonparametric permutation tests (Churchill and Doerge, 1994) based on the method of Westfall and Young (1993) to determine the FWER for the data from each year.

Multiple regression was used to develop a model that best explained the underlying phenotypic variation using the REG procedure of SAS with the stepwise selection option. All alleles identified as having an association with the trait of interest were initially included in the model, and only those alleles that remained significant at $\alpha = 0.05$ were retained.

RESULTS AND DISCUSSION

Genetic Linkage Mapping

The genetic map consists of eight consensus LGs representing the eight basic chromosomes of the alfalfa genome (Fig. 1) and covers 546 cM. The size of the basic chromsome complement of alfalfa (i.e., $x = 8$) is 0.856 pg DNA (Blondon et al., 1994), or approximately 840 Mbp. Therefore, the length of the map represents approximately 1.54 Mbp cM^{-1}. The consensus parental maps were decomposed into 32 cosegregation groups for each parental genome (Supplementary Fig. 1; an example of cosegregation groups corresponding to LG 7 are shown in Fig. 2). Determining cosegregation groups provides a more precise view of linkage relationships among marker alleles and enables easier localization of potential QTL positions.

Thirty-two percent of the alleles exhibited segregation distortion (Supplementary Table 2). Compared to the three previous tetraploid alfalfa mapping populations, this level of distortion was similar to that found by Julier et al. (2003), but higher than the results of Brouwer and Osborn (1999) and Sledge et al. (2005). The Brouwer and Osborn (1999) and Sledge et al. (2005) mapping populations consisted almost entirely of single-dose restriction fragments (single-dose alleles), which by definition segregate in a 1:1 present-to-absent ratio (Wu et al., 1992). In comparison, Julier et al. (2003) and our study included a large number of alleles segregating at higher-order ratios, which likely resulted in increased levels of segregation

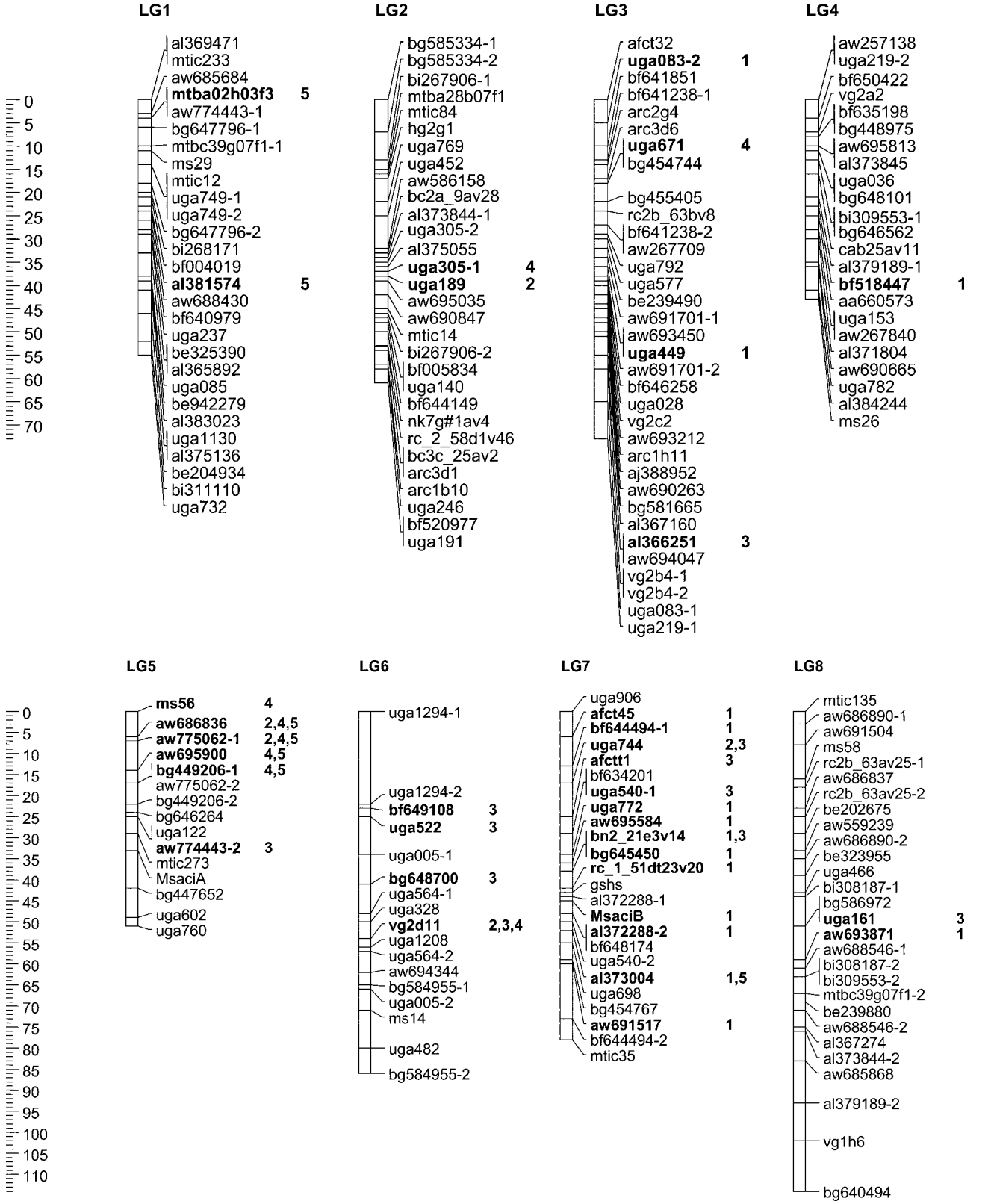

Fig. 1. Consenus genetic linkage maps of F_1 progeny of the cross between WISFAL-6 and ABI408. Loci containing alleles associated with biomass production are in boldface and noted by the environment in which they were identified (1, IA1999; 2, IA2000; 3, IA2001; 4, NY2000; and 5, NY2001).

distortion compared to the exclusive use of single-dose alleles. This population did not exhibit any deformed or otherwise mutant plants as is often seen in diploid populations derived from selfing (e.g., Brummer et al., 1993). Further, the 200 individuals used in this experiment were chosen by simply potting the first 200 seedlings

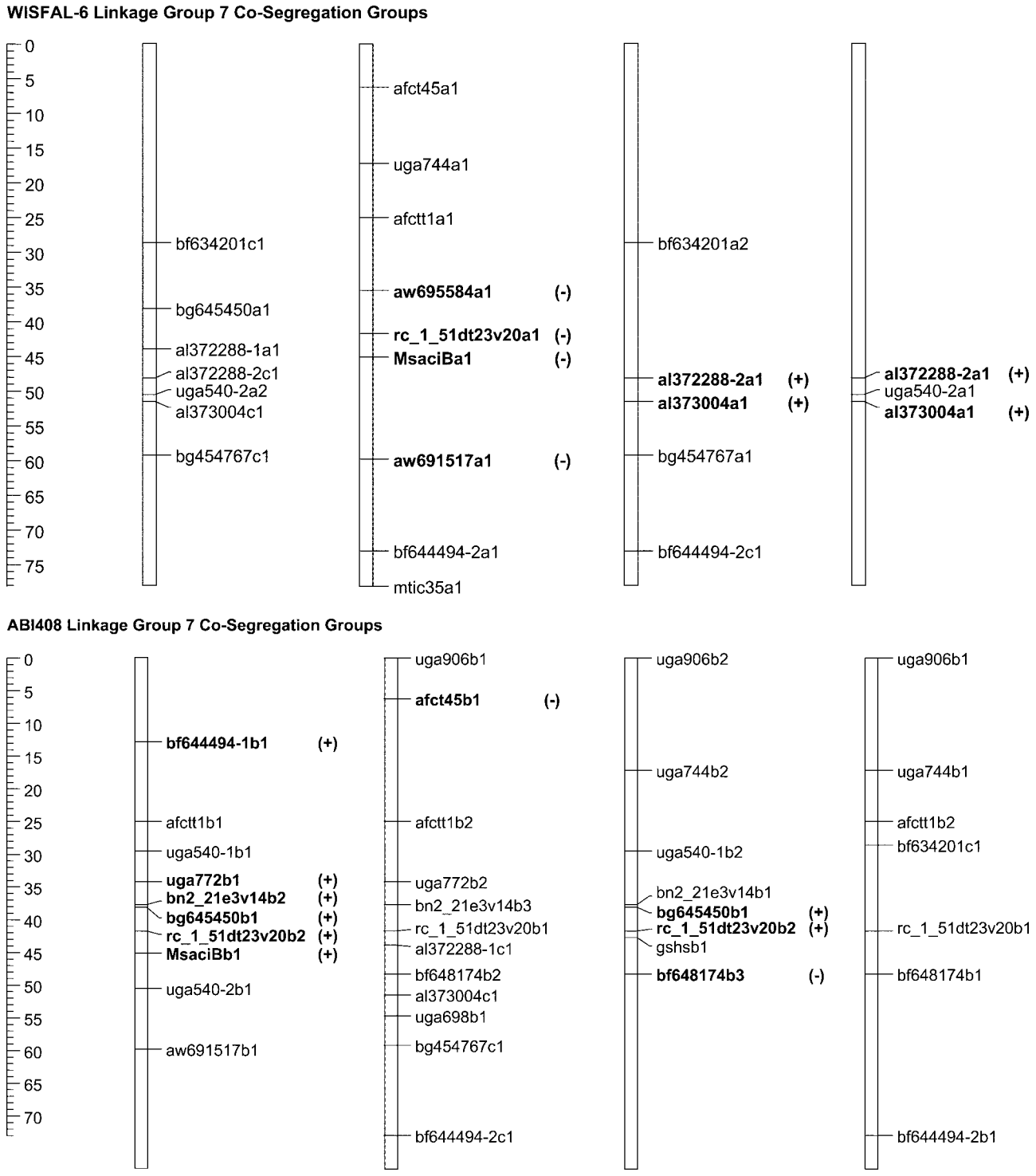

Fig. 2. Location of alleles associated with biomass production (IA 99) on the LG 7 cosegregation groups of each parent. Alleles with positive effects are marked with a plus and alleles with negative effects are marked with a minus.

regardless of their vigor. They were identified before cloning for the field phase, so differential ability to make cuttings could also not have caused the distortion. Of course, selection at the zygote stage could have been operative in this cross. Distortion did not appear to have any directionality favoring one parent or the other.

We identified 25 putatively duplicated loci (18 SSR; 7 RFLP). With six exceptions, these loci were duplicated on the same LG. Some of the duplications may have resulted from inaccurate designations among alleles.

Biomass Production

Location and genotype-by-location interaction (GxL) variances were computed from the IA data (Table 1). The location effect was only present in 2000, and al-

Table 1. Variance component estimates and their standard errors for genotype (σ^2_G), genotype-by-environment (σ^2_{GL}), and error (σ^2_E) and broad-sense heritability estimates on an entry-mean ($H^2_{(EntryMean)}$) and a plot basis ($H^2_{(Plot)}$) for each year averaged across two Iowa locations and for Ithaca, NY.

	IA99	IA00	IA01	NY00	NY01
σ^2_G	2442 ± 289†	2547 ± 408	2450 ± 446	229 ± 28	983 ± 121
σ^2_{GL}	457 ± 84	621 ± 274	406 ± 360	N/A	N/A
σ^2_E	1443 ± 60	7859 ± 328	9910 ± 456	194 ± 11	816 ± 48
$H^2_{(Plot)}$	0.56 ± 0.03	0.23 ± 0.03	0.19 ± 0.03	0.54 ± 0.04	0.55 ± 0.04
$H^2_{(Entry\ Mean)}$	0.86 ± 0.02	0.66 ± 0.05	0.63 ± 0.06	0.83 ± 0.02	0.83 ± 0.02

† ± Standard errors of estimates.

though GxL variation was present in 1999 and 2000, it was typically an order of magnitude smaller than the variance associated with genotype. Subsequent analyses of biomass production and QTL identification were based on data combined across the two IA locations and separately for NY. Therefore, five location–year combinations (termed "environments" in the subsequent discussion) were identified (IA99, IA00, IA01, NY00, and NY01).

The variance due to genotype was often larger than the variance associated with error, except in IA00 and IA01, when machine harvest resulted in lower precision (Table 1). The broad-sense heritability estimates, based on individual plots and on entry means, indicated that biomass production is under substantial genetic control in this population. Even in IA00 and IA01, the entry mean heritabilities were high. Obviously, for applied breeding purposes, estimates of narrow-sense heritability are of more importance than broad-sense estimates, but the structure of our population did not allow this estimation. Narrow-sense heritabilities would be smaller, and possibly considerably smaller, than the broad-sense estimates.

Biomass production ranged widely within the population. Transgressive segregation in both directions was present at all locations and years (Table 2) and was consistently associated with the same genotypes across years and within locations (data not shown). The parents did not differ for biomass production in any location or year (Table 2). The mean performance of the F_1 population was not higher than either the high-parent (ABI408 had a numerically higher yield in most environments) or the mid-parent value (Table 2). The ability to discriminate among genotypes declined across years as the error associated with biomass determination escalated due to plant mortality and, in IA, mechanical harvesting. The increase in error variance is not uncommon for aging forage trials.

Table 2. Mean forage biomass production per year of the F_1 population, its parents, and the high- and low-yielding F_1 genotypes averaged across two Iowa locations and in Ithaca, NY.

	IA99	IA00	IA01	NY00	NY01
	g plant^{-1}				
F_1 population mean	167 ± 27†	357 ± 62	285 ± 71	46 ± 14	111 ± 30
ABI408 mean	140	337	212	41	100
WISFAL-6 mean	129	333	219	31	81
High F_1 genotype	295	467	420	110	205
Low F_1 genotype	27	177	97	9	15

† ± Standard errors of estimates.

Phenotypic correlations of biomass production between the years in IA were moderate, while the phenotypic correlation between the two years at NY was high (Table 3). Genetic correlations between years were all high in both IA and NY (0.57 to 0.95). These results suggest that biomass production has a similar genetic basis from year to year, both under different environmental conditions and at different developmental stages of the plants. In addition, although the differing experimental designs and missing first harvest data from the NY location precluded the combined analysis of the data, simple correlations between IA and NY based on the years after establishment (IA99 and NY00; IA00 and NY01) were ~0.60 for both years; the correlation between IA00 and NY00 was ~0.6 and between IA01 and NY01 ~0.4 (all with p values < 0.0001). The yield data were normally distributed with no obvious outlier points that could have unduly affected the correlations (data not shown).

Marker–Phenotype Associations

Because we scored all segregating alleles, some of which had complex segregation ratios, we could not apply traditional QTL mapping techniques widely used in diploid populations. Interval mapping can be conducted on polyploids if maps are based solely on single-dose alleles, which is not the case in our experiment. We opted to use single-marker analysis even though it cannot precisely localize QTLs or determine the number of QTLs in a given region due to confounding of recombination and genotypic value (Bernardo, 2002). However, single-marker analysis is a proven method that works well for initial identification of QTLs and has been previously used in alfalfa mapping studies (Brouwer et al., 2000; Sledge et al., 2002). More comprehensive statistical genetic theory for interval mapping in auto-

Table 3. Phenotypic (top) and genetic (bottom) correlations ± standard errors between total yearly biomass production from 3 yr in Iowa averaged across two locations and from 2 yr in Ithaca, NY.

	Correlation	
	IA00	IA01
IA99	0.35 ± 0.03 0.78 ± 0.05	0.26 ± 0.03 0.57 ± 0.07
IA00		0.36 ± 0.03 0.90 ± 0.06
	NY01	
NY00	0.74 ± 0.02 0.95 ± 0.02	

polyploids is being developed (Hackett et al., 2001; Cao et al., 2005).

Forty-one alleles were associated with biomass production in at least one environment (Table 4). Seven of these alleles were identified in more than one year or location, providing further support that they are actually linked to QTLs and not associated with biomass yield due to random chance. In addition, a stringent permutation test based on a FWER set to control false positive results at the 0.05 probability level also identified several marker alleles associated with biomass production (Table 4). Alleles that were associated with biomass production on a given LG were often linked (e.g., LG 7). Different alleles at the same locus were associated with biomass in different environments, as in the case of bn2_21e3v14, in which one allele had a positive effect in IA99 but another allele a negative effect in IA01. Further, in some cases, the same locus had alleles associated with biomass yield in each parent (e.g., *MsaciB*); this would appear to be unlikely to have arisen simply due to chance. The gene *MsaciB* is known to be related to winter hardiness (Monroy et al., 1993), and its association with biomass production is interesting and may suggest that biomass yield is affected by winter injury. Nevertheless, some of the alleles we identified may be falsely associated with biomass production.

Each consensus LG contributed alleles that were associated with biomass production in at least one environment. Both parents contributed loci with both positive (e.g., *ms56*) and negative (e.g., *vg2d11b1*) effects on biomass production, often contributing positive and negative associations from the same LG. On LG 7, WISFAL-6 contributed six alleles associated with biomass production. The four alleles with negative effects are located on one cosegregation group, and the two alleles positively associated with biomass production are located on other groups (Fig. 2). Thus, we show that different homologous chromosomes within the same plant can contribute alleles with both positive and negative effects on a trait.

One method to identify QTL by environment interactions is to examine trends of marker–phenotype associations across environments (Lynch and Walsh, 1998). Of the seven alleles associated with biomass yield in

Table 4. Phenotypic effects (g plant^{-1}) of alleles associated with biomass production, based on single-marker analysis ($\alpha < 0.01$).

Marker allele†	Linkage group	IA 1999	IA 2000	IA 2001	NY 2000	NY 2001
		g plant^{-1}				
mtba02h03f3a3	1					+15*
al381574c3	1					+14*
uga189a3	2		−27**			
uga305–1b1	2				+7**	
uga671a1	3				+9**	
uga449b2	3	−21*				
uga083–2b2	3	−27**				
al366251c1	3			−29**		
bf518447b1	4	+21*				
aw695900a1	5				+9**‡	+17**
aw744443–2a4	5			−31*		
bg449206–1a3	5				+12**	+19*
ms56a1	5				+7**	
ms56a2	5				−9**	
aw686836c3	5		+28**		+10**‡	+18**§
aw775062–1c1	5		+33**		+10**	+16*
bf649108b3	6			+32**		
uga522b2	6			−25*		
vg2d11b1	6		−22*		−7**	
vg2d11b2	6			+28**		
bg648700c2	6			+27*		
al372288–2a1	7	+29**				
aw691517a1	7	−26**				
MsaciBa1	7	−31**‡				
rc_1_51dt23v20a1	7	−27**				
al373004a1	7	+30**				+17*
aw695584a1	7	−23**				
afct45b1	7	−21**				
afctt1b2	7			−31*		
bf644494–1b1	7	+30**				
bg645450b1	7	−28**				
bn2_21e3v14b1	7			−31**		
bn2_21e3v14b2	7	+33**‡				
MsaciBb1	7	+20*				
rc_1_51dt23v20b2	7	+25**				
uga540–1b2	7			−26**		
uga744b1	7			+27*		
uga744b2	7		−24**	−26**		
uga772b1	7	+29**§				
uga161b1	8			+26**		
aw693871c3	8	+41**				

* Marker–trait association is significant at the 0.01 probability level.
** Marker–trait association is significant at the 0.005 probability level.
† Allele designations following the locus name are as follows: a, WISFAL-6; b, ABI408; c, both parents.
‡ Familywise error rate significant at 0.05 level.
§ Familywise error rate significant at 0.10 level.

more than one year, only two, *aw775062–1c1* and *aw686836c1*, were identified in more than two environments; none was present in all environments (Table 4). The remaining 34 alleles exhibited an association with biomass production in only one environment. Although some of these may be false positive associations, interaction of alleles with years is also a probable explanation, particularly because several of these alleles were also identified by the conservative permutation test (Table 4) and by multiple regression analyses (Table 5). The changing environmental conditions and developmental trajectory of the plants from year to year would likely impact the genetic control of biomass production. The subset of alleles identified in more than one environment indicates that at least some genomic regions are important for biomass production across various environmental conditions, in agreement with similar studies in tree species (Wu, 1998; Lerceteau et al., 2001). Possibly due to the change in harvest management in IA and/or to aging of plants at both locations, our ability to detect allelic associations with biomass QTLs declined over years, and this may be a reason that some alleles identified in one year were not detected otherwise.

Multiple regression models from the different environments explained between 13 and 36% of the phenotypic variation associated with biomass production (Table 5). These are likely overestimates of the actual amount of genetic variation controlled by these loci due to the upward bias in estimation of QTL effects inherent in this type of study (Utz et al., 2000), but it also indicates that potential improvement from capitalizing on these alleles may be substantial.

Application of Results to Crop Breeding

The results presented here show that the *M. sativa* subsp. *falcata* genotype WISFAL-6 contains QTL alleles that increase biomass production; we can surmise that other *M. sativa* subsp. *falcata* germplasm, including that which does not produce biomass yield heterosis in interspecific progeny (Riday and Brummer, 2005), may also harbor beneficial alleles. In this experiment, LG 7 contained several alleles associated with biomass production in most of the environments. A previous genetic map constructed from an F_2 diploid alfalfa population (Brummer et al., 1993) expressed very high levels of segregation distortion on LG 7 (which corresponds to group 4 on that map), with progeny ratios skewed toward excess heterozygotes, reaching 90% heterozygous progeny for some loci. The fact that LG 7 has a striking effect on fitness (biomass productivity) in a tetraploid population further supports the hypothesis that this region may be involved with heterosis for biomass productivity. Our result suggests a connection with heterosis in that positive alleles from both germplasm sources are present in the highest-yielding hybrid progeny. While the identification of heterosis in a single full sib population is confounded with yield per se, this result bears further investigation through studies in other populations.

A move toward marker-assisted selection in *M. sativa* faces several obstacles. Little linkage disequilibrium may be present in modern alfalfa breeding populations that have had the potential for substantial amounts of recombination during repeated rounds of recurrent selection (Williams, 1998; Kidwell et al., 1999). However, the identification of several significant marker–phenotype associations in this experiment is heartening, and the important associations on LG 5 and 7 are obvious targets for fine mapping. Our analysis of the genetic determinants of biomass production in tetraploid alfalfa offers the first steps toward the use of molecular markers in an alfalfa recurrent selection program. However, to be useful in a selection program, associations would require improved localization and the identification of a tight linkage between marker and trait or, ideally, the identification of a functional marker that defines the beneficial allele (Andersen and Lübberstedt, 2003). Another concern is that the amount of linkage disequilibrium in an F_1 population makes isolating the QTLs to a small interval difficult. Further exploration of biomass QTLs in other populations and environments is warranted to substantiate these results.

Table 5. Alleles forming best-fit models based on stepwise multiple regression for biomass production and their partial R^2 values from each year in Iowa and in Ithaca, NY.

Marker allele†	Linkage group	IA99	IA00	IA01	NY00	NY01
mtba02h03f3a3	1					0.05
uga189a3	2		0.08			
uga305–1b1	2				0.06	
uga671a1	3				0.04	
bg449206–1a3	5					0.04
aw686836c3	5				0.12	0.11
aw775062–1c1	5		0.04		0.03	
uga522b2	6			0.04		
vg2d11b1	6		0.05			
al372288–2a1	7	0.04				
bf644494–1b1	7	0.16				
MsaciBa1	7	0.12				
bg645450b1	7	0.04				
uga744b2	7		0.03	0.09		
Cumulative R^2		0.36	0.20	0.13	0.25	0.19
Total g plant^{-1}‡		111	106	51	36	52

† Allele designations following the locus name are as follows: a, WISFAL-6; b, ABI408; c, both parents.
‡ Total g/plant represents the sum of the yield differentials between plants with and without positive alleles for yield at each of the marker loci identified for a particular environment.

Perspectives

This study is the first of its kind and is necessarily exploratory. Although it presents evidence that QTLs associated with biomass production can be identified, even in complex tetraploid populations, the experiment has some limitations. First, we used clonally propagated, spaced plants. This obviously differs from a commercial field, both in mode of establishment and in intraplant competition. Thus, further research done with progeny families grown in swards or semiswards would be very useful. Second, we are only evaluating two parental genotypes, a restricted germplasm base not commonly used in alfalfa breeding. Moving to population-based methods of mapping, such as association analysis (e.g., Skøt et al., 2005), would help alleviate some of that concern. Additionally, mapping in populations with more recombination, and using a higher density of markers, particularly ones for which allele dosages can be obtained, would make the localization of QTLs more tractable. Finally, we are conducting a more detailed look at harvest-by-harvest biomass production, and further work on that area, coupled with analyses to separate the effects of plant age from environment, should be undertaken.

ACKNOWLEDGMENTS

We thank Drs. Dan Nettleton, Heathcliffe Riday, Meenakshi Santra, M. Maroof Shah, Cara Lea Council, and Ryan Kunz for assistance. This research was supported by USDA-NRI (97-35300-4573), USDA-IFAFS (00-52100-9611), and Hatch Regional Research Project NE-1010 (all to E.C.B.) and an Iowa State University Plant Sciences Institute Fellowship (to J.G.R.).

REFERENCES

Andersen, J.R., and T. Lübberstedt. 2003. Functional markers in plants. Trends Plant Sci. 8:554–560.

Austin, D.F., and M. Lee. 1998. Detection of quantitative trait loci for grain yield and yield components in maize across generations in stress and nonstress environments. Crop Sci. 38:1296–1308.

Bernacchi, D., T. Beck-Bunn, Y. Eshed, S. Inai, J. Lopez, V. Petiard, J. Uhlig, D. Zamir, and S.D. Tanksley. 1998. Advanced backcross QTL analysis in tomato: I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. Theor. Appl. Genet. 97:381–397.

Bernardo, R. 2002. Breeding for quantitative traits in plants. Stemma Press, Woodbury, MN.

Bingham, E.T. 1993. Registration of WISFAL alfalfa (*Medicago sativa* ssp. *falcata*) tetraploid germplasm derived from diploids. Crop Sci. 33:217–218.

Blondon, F., D. Marie, and S. Brown. 1994. Genome size and base composition in *Medicago sativa* and *M. truncatula* species. Genome 37:264–270.

Brouwer, D.J., S.H. Duke, and T.C. Osborn. 2000. Mapping genetic factors associated with winter hardiness, fall growth, and freezing injury in autotetraploid alfalfa. Crop Sci. 40:1387–1396.

Brouwer, D.J., and T.C. Osborn. 1999. A molecular marker linkage map of tetraploid alfalfa (*Medicago sativa* L.). Theor. Appl. Genet. 99:1194–1200.

Brummer, E.C. 1999. Capturing heterosis in forage crop cultivar development. Crop Sci. 39:943–954.

Brummer, E.C., J.H. Bouton, and G. Kochert. 1993. Development of an RFLP map in diploid alfalfa. Theor. Appl. Genet. 86:329–332.

Brummer, E.C., M.M. Shah, and D. Luth. 2000. Reexamining the relationship between fall dormancy and winter hardiness in alfalfa. Crop Sci. 40:971–977.

Cao, D., B.A. Craig, and R.W. Doerge. 2005. A model selection-based interval-mapping method for autopolyploids. Genetics 169: 2371–2382.

Choi, H.-K., D. Kim, T. Uhm, E. Limpens, H. Lim, J.-H. Mun, P. Kaló, R.V. Penmetsa, A. Seres, O. Kulikova, B.A. Roe, T. Bisseling, G.B. Kiss, and D.R. Cook. 2004. A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166:1463–1502.

Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963–971.

DeWinton, D., and J.B.S. Haldane. 1931. Linkage in the tetraploid *Primula Sinensis*. J. Genet. 24:121–144.

Diwan, N., J.H. Bouton, G. Kochert, and P.B. Cregan. 2000. Mapping of simple sequence repeat (SSR) DNA markers in diploid and tetraploid alfalfa. Theor. Appl. Genet. 101:165–172.

Diwan, N., P. Cregan, G. Bauchan, and A. Bhagwat. 1997. Simple sequence repeat (SSR) DNA markers in alfalfa and perennial and annual *Medicago* species. Genome 40:887–895.

Doyle, J.J., and J.I. Doyle. 1990. Isolation of plant DNA from fresh tissue. BRL Life Technologies: Focus 12:13–15.

Echt, C.S., K.K. Kidwell, S.J. Knapp, T.C. Osborn, and T.J. McCoy. 1994. Linkage mapping in diploid alfalfa (*Medicago sativa*). Genome 37:61–71.

Eujayl, I., M.K. Sledge, L. Wang, G.D. May, K. Chekhovskiy, J.C. Zwonitzer, and M.A.R. Mian. 2003. *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for Medicago spp. Theor. Appl. Genet. 108:414–421.

Galitski, T., A.J. Saldanha, C.A. Styles, E.S. Lander, and G.R. Fink. 1999. Ploidy regulation of gene expression. Science 285:251–254.

Groose, R.W., W.P. Kojis, and E.T. Bingham. 1988. Combining ability differences between isogenic diploid and tetraploid alfalfa. Crop Sci. 28:7–10.

Hackett, C.A., J.E. Bradshaw, and J.W. McNicol. 2001. Interval mapping of quantitative trait loci in autotetraploid species. Genetics 159:1819–1832.

Hackett, C.A., and Z.W. Luo. 2003. TetraploidMap: Construction of a linkage map in autotetraploid species. J. Hered. 94:358–359.

Haldane, J.B.S. 1930. Theoretical genetics of autoploids. J. Genet. 22: 359–372.

Holland, J.B. 2006. Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation in SAS Proc MIXED. Crop Sci. 46:642–654.

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martinez. 2003. Estimating and interpreting heritability for plant breeding. Plant Breed. Rev. 22:9–111.

Julier, B., S. Flajoulot, P. Barre, G. Cardinet, S. Santoni, T. Huguet, and C. Huyghe. 2003. Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. BMC Plant Biol. 3:9.

Kidwell, K.K., L.M. Hartweck, B.S. Yandell, P.M. Crump, J.E. Brummer, J. Moutray, and T.C. Osborn. 1999. Forage yields of alfalfa populations derived from parents selected on the basis of molecular marker diversity. Crop Sci. 39:223–227.

Kiss, G.B., G. Csanádi, K. Kálmán, P. Kaló, and L. Ökrész. 1993. Construction of a basic genetic map for alfalfa using RFLP, RAPD, isozyme and morphological markers. Mol. Gen. Genet. 238:129–137.

Lamb, J.F.S., C.C. Sheaffer, L.H. Rhodes, R.M. Sulc, D.J. Undersander, and E.C. Brummer. 2006. Five decades of alfalfa cultivar improvement: Impact on forage yield, persistence, and nutritive value. Crop Sci. 46:902–909.

Lerceteau, E., A.E. Szmidt, and B. Andersson. 2001. Detection of quantitative trait loci in *Pinus sylvestris* L. across years. Euphytica 121:117–122.

Littell, R.C., G.A. Milliken, W.W. Stroup, and R.D. Wolfinger. 1996. SAS® system for mixed models. SAS Institute, Cary, NC.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, Sutherland, MA.

Mather, K. 1936. Segregation and linkage in autotetraploids. J. Genet. 32:287–314.

Monroy, F.A., Y. Castonguay, S. Laberge, F. Sarhan, L.-P. Vézina, and R.S. Dhindsa. 1993. A new cold-induced alfalfa gene is associated with enhanced hardening at subzero temperature. Plant Physiol. 102:873–879.

Riday, H., and E.C. Brummer. 2002. Forage yield heterosis in alfalfa. Crop Sci. 42:716–723.

Riday, H., and E.C. Brummer. 2005. Heterosis in a broad range of alfalfa germplasm. Crop Sci. 45:8–17.

Rumbaugh, M.D., J.L. Caddel, and D.E. Rowe. 1988. Breeding and quantitative genetics. *In* A.A. Hanson et al. (ed.) Alfalfa and alfalfa improvement. Agron. Monogr. 29. ASA, CSSA, and SSSA, Madison, WI.

Skøt, L., M.O. Humphreys, I. Armstead, S. Heywood, K.P. Skot, R. Sanderson, I.D. Thomas, K.H. Chorlton, and N.R.S. Hamilton. 2005. An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). Mol. Breed. 15:233–245.

Sledge, M.K., J.H. Bouton, M. Dall'Agnoll, W.A. Parrott, and G. Kochert. 2002. Identification and confirmation of aluminum tolerance QTL in diploid *Medicago sativa* subsp. *coerulea*. Crop Sci. 42:1121–1128.

Sledge, M.K., I.M. Ray, and G. Jiang. 2005. An expressed sequence tag SSR map of tetraploid alfalfa (*Medicago sativa* L.). Theor. Appl. Genet. 111:980–992.

Thoquet, P., M. Ghérardi, E.-P. Journet, A. Kereszt, J.-M. Ané, J.-M. Prosperi, and T. Hughet. 2002. The molecular genetic linkage map of the model legume *Medicago truncatula*: An essential tool for comparative legume genomics and the isolation of agronomically important genes. BMC Plant Biol. 2:1.

Utz, H.F., M.E. Melchinger, and C.C. Schön. 2000. Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154:1839–1849.

Van Ooijen, J.W., and R.E. Voorrips. 2001. JoinMap 3.0: Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, Netherlands.

Voorrips, R.E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77–78.

Westfall, P.H., and S.S. Young. 1993. Resampling-based multiple testing: Examples and methods for *p*-value adjustment. John Wiley & Sons, New York.

Williams, C.G. 1998. QTL mapping in outbred pedigrees. *In* A.H. Paterson (ed.) Molecular dissection of complex traits. CRC Press, Boca Raton, FL.

Wu, R.L. 1998. Genetic mapping of QTLs affecting tree growth and architecture in *Populus*: Implication for ideotype breeding. Theor. Appl. Genet. 96:447–457.

Wu, K.K., W. Burnquist, M.E. Sorrells, T.L. Tew, P.H. Moore, and S.D. Tanksley. 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. 83:294–300.